DOCUMENT RESUME

ED 144 533 IR 004 896

AUTHOR Clements, David T.
TITLE Voice Output for Student Information Inquiry.
PUB DATE 75
NOTE 18p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement; Cocurricular Activities; *Computer Oriented Programs; Cost Effectiveness; *Data Bases; *Information Networks; *Input Output Devices; Junior Colleges; Student Characteristics; *Student Records; Use Studies
IDENTIFIERS *Voice Synthesizers

ABSTRACT

This paper describes the application of a touch-tone/Synthesized Voice Output facility for the administrative inquiry of a student information data base. It contrasts the traditional audio response with the synthesized voice output and describes the principles of voice synthesizer operation and the software resources required. It also addresses concerns on the privacy of information, the need for message brevity, and the repeatability of messages. Flow charts are used throughout to depict the logic of the system and its user interface. Preliminary conclusions indicate: (1) a favorable user response due to the ease of access and the availability of "terminals," (2) a limited scope in application when considering that complex information is best communicated in printed form, and (3) a favorable cost to user ratio. (DAG)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

ED144533

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# VOICE OUTPUT FOR STUDENT INFORMATION INQUIRY

David T. Clements

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

David T. Clements

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

IR004896

# VOICE OUTPUT FOR STUDENT INFORMATION INQUIRY

David T. Clements, Senior Programmer/Analyst
District Student Information Systems
Coast Community College District
Costa Mesa, California 92626

The Coast Community College district is currently implementing several Synthesized Voice Output applications in both their Computer Aided Instruction (CAI) and On-line Student Information Systems. This paper focuses mainly on the efforts expended in providing a touch-tone/voice-output facility for Administrative inquiry of the Student Information Data Base. The author's intent is to familiarize the reader with the components of the facility and to propose its employment as an extremely effective medium from both cost and user points of view.

## 1. INTRODUCTION

Man has always regarded his ability to speak as an endowment which elevates him from all other earthly creatures. Whether he is now ready to include the "monstrous" digital computer into his elite status is debatable. But the technology of Audio Response has grown tremendously in the recent years and it has proven to be, in many applications ranging from Credit Inquiry to CAI for the blind [3], an ideal Man/Machine Communication.

A recent research article [5] on Voice Response listed some 18 American suppliers of speech generating equipment and systems. The available hardware appears to fall into two categories which, for the sake of delineation, might be termed Traditional Audio Response and Synthesized Voice Output. The major differences between these two types is shown in figure 1. The Traditional Audio Response unit contains sophisticated storage systems which are "loaded" with pre-recorded syllables, words, phrases or perhaps full paragraphs. Each of these "sound units" is addressable. The host program sends a request for a particular unit and it is output (or more properly, "played").

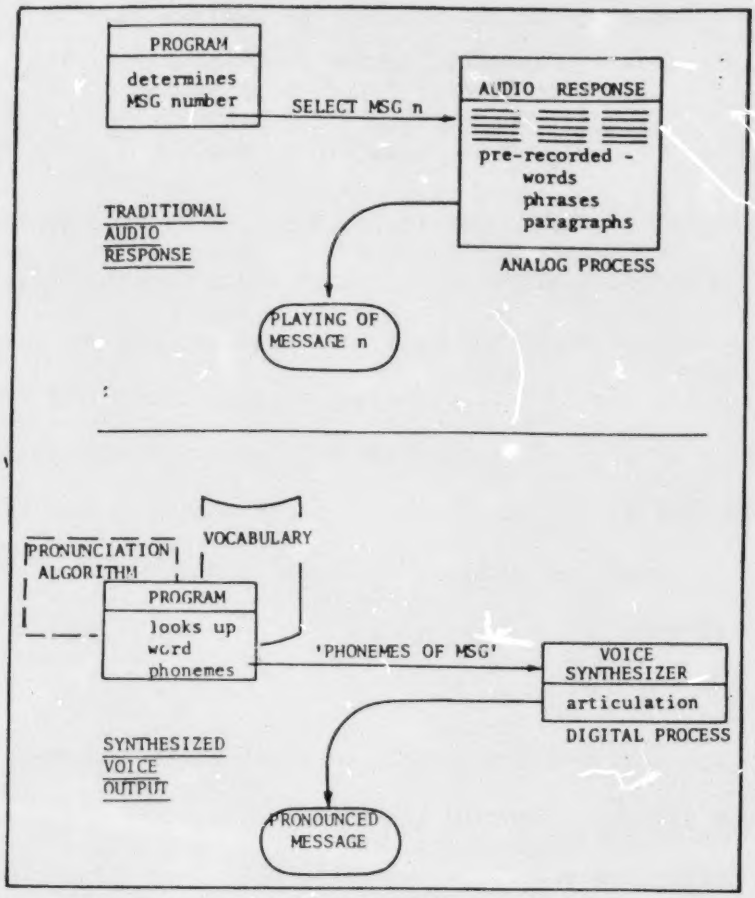


figure 1: Pre-recorded Audio Response versus Voice Synthesis

With the exception of some control unit logic, it is a total analog procedure.

A Voice Synthesizer differs from the Traditional Audio Response systems in that there are no pre-stored, pre-recorded sounds. It reacts only to digital excitement. The host program determines the phonetic structure of its output by means of a vocabulary lookup procedure or pronunciation algorithm and sends a binary string of appropriate pronunciation codes to the device for articulation.

We decided to employ a Voice Synthesizer* rather than a traditional Audio Response system for several reasons. Most notably because it seemed ideal for

*The district utilizes a VOTRAX-VI™ voice synthesizer, supplied by the Vocal Interface Division of the Federal Screw Works, Detroit, Michigan.

a wide range of applications. Within our CAI systems, there are many possible uses starting with the study of speech sounds themselves on up to a full vocal interface for blind students, similar to the work in progress at Michigan State University. The Administrative side of the house also could visualize several applications which shall be summarized later in this paper.

Another attractive quality of Synthesized Voice is that there is no limit on vocabulary size such as exists with those devices requiring pre-recorded units. Vocabulary tables may be stored easily on disk. Our structure will accommodate in the area of 8000-10000 words with their phonetic codings per one 3330 disk cylinder. Also, an interactive procedure may be quickly developed to add or refine the vocabulary as it is being accessed by other tasks.

We also considered the ability to dynamically create sentences within the host program a valuable asset. That is, constant text may be spliced with variable data to create a meaningful output string.

The most convincing argument in favor of Voice Synthesis is its extreme economy. Our synthesizer may be purchased for approximately $3500. This is well under the cost of the other Audio Response units comparable to IBM's 7770 which starts at $60,000 and go as high as $230,000.*

---

*The amounts quoted are per reference [2], specifically IBM's 7770. In all fairness, there are vendors which approach 7770 compatibility at a more favorable price. There are also several vendors supplying devices with limited vocabularies for specific application areas and their prices in some cases are significantly less. To the author's knowledge, however, there is currently only one supplier of pure Voice Synthesis equipment (as described in this paper); the Vocal Interface Division of the Federal Screw Works.

## 2. PRINCIPLES OF VOICE SYNTHESIZER OPERATION AND ITS REQUIRED RESOURCE

Phonetic coding for a Voice Synthesizer is shown in figure 2. Our synthesizer currently reacts to some 60 phonemes. (A phoneme is the smallest unit of speech.) There are also pausing and inflection codes. Note that the phonemes and pauses only use the low order six positions of an eight bit configuration. Inflection codes utilize the high order two bit structure.

▶ THERE ARE SOME 60 PHONEMES,

| PHONEME NAME | BINARY REPRESENTATION |
|---|---|
| AW (as in law) | 00111101 |
| AH (as in car) | 00100100 |
| AY (as in eight) | 00100001 |
| AE (as in cat) | 00101110 |
| B (as in book) | 00001110 |
| etc.... | .... |
| etc.... | .... |

▶ PAUSE CODES,

| PAUSE NAME | BINARY REPRESENTATION |
|---|---|
| PAØ (short pause) | 00000011 |
| PA1 (med. pause) | 00111110 |
| PA2 (long pause) | 00110000 |

▶ AND INFLECTION CODES which are OR'D into bits Ø and 1

| INFLECTION | BINARY REPRESENTATION |
|---|---|
| IN1 (low) | 10000000 |
| IN2 (normal) | 11000000 |
| IN3 (high) | 00000000 |
| IN4 (higher) | 01000000 |

▶ EXAMPLE: The word HELP is constructed as follows-

| inflection / phoneme | 3 / H | 1 / EH1 | 1 / UH3 | 1 / L | 1 / P |
|---|---|---|---|---|---|
| =binary | 00011011 | 10000001 | 10100011 | 10011000 | 10100101 |

figure 2: Phonetic coding and its binary representation

An inflection is properly OR'd to a phoneme for the desired sound. The example shown for pronouncing the word HELP indicates inflection/phoneme notation and the resulting binary codes. The H sound, of course, must be stressed; thus a level 3, or high inflection. Note also that two phonemes are required to accomplish the diphthong vowel sound.

The device obviously belongs in a real-time environment. Because it has been engineered to interface per the EIA standard RS-232, there are several configuration opportunities. We have supported two different connections. Figure 3 depicts the synthesizer connected in parallel with a CAI terminal. This is accomplished by special strapping and subsetting the RS-232 interface [2]. What appears on the terminal is also spoken; an ideal coordination for trying new words and inflections.

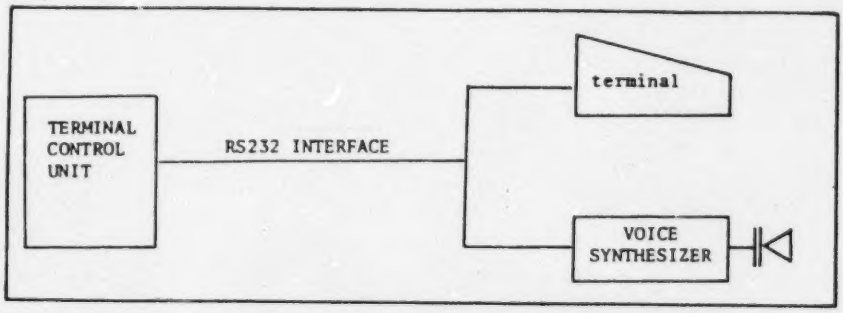


figure 3: Voice synthesizer is parallel with terminal

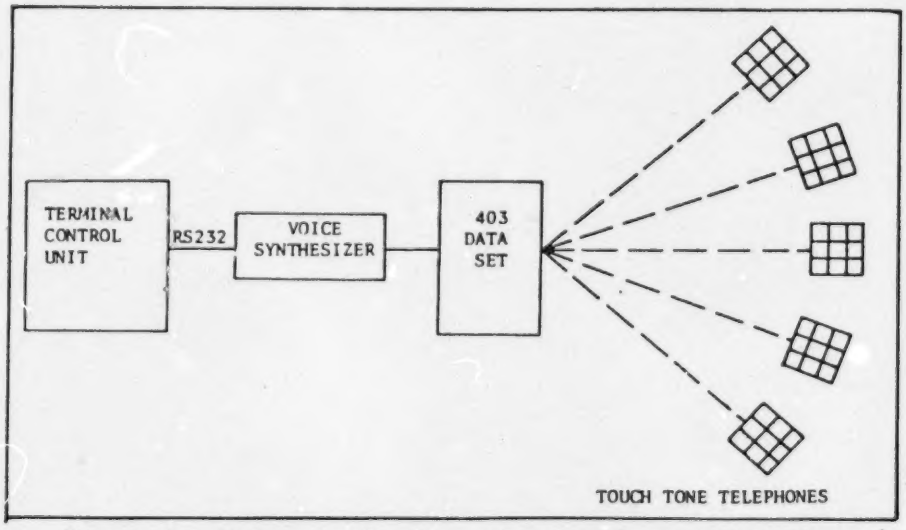


figure 4: Voice synthesizer with touch-tone interface

Our other configuration shown in figure 4 interfaces the synthesizer to a Bell 403E6 data set and subsequent acoustical touch-tone interaction.*

*Unfortunately only one touch-tone telephone may interact with the configuration at a time. We are hopeful this limitation will be soon alleviated by an economic multiplexing option.

## 3. SOFTWARE PROCEDURES FOR CONTROLLING VOICE SYNTHESIZERS

The Coast Community College district supports two teleprocessing systems; namely, APL/SV for CAI activities and ENVIRON/1 for Administrative on-line applications.* Speech synthesis is used in both systems (ref. figure 5). Currently, we develop new vocabulary in an interactive mode under APL/SV. The vocabulary table (Lexicon) is stored in a sequential file. Of course, small vocabularies can also be stored as direct variables in the APL work-space. The APL vocabulary table is subsequently read by a batch PL/1 program which hashes the elements into a new structure called PHAST (Phonetic Hash Table). Administrative on-line applications may then access the vocabulary in ENVIRON/1 thru a common Lookup Procedure module.

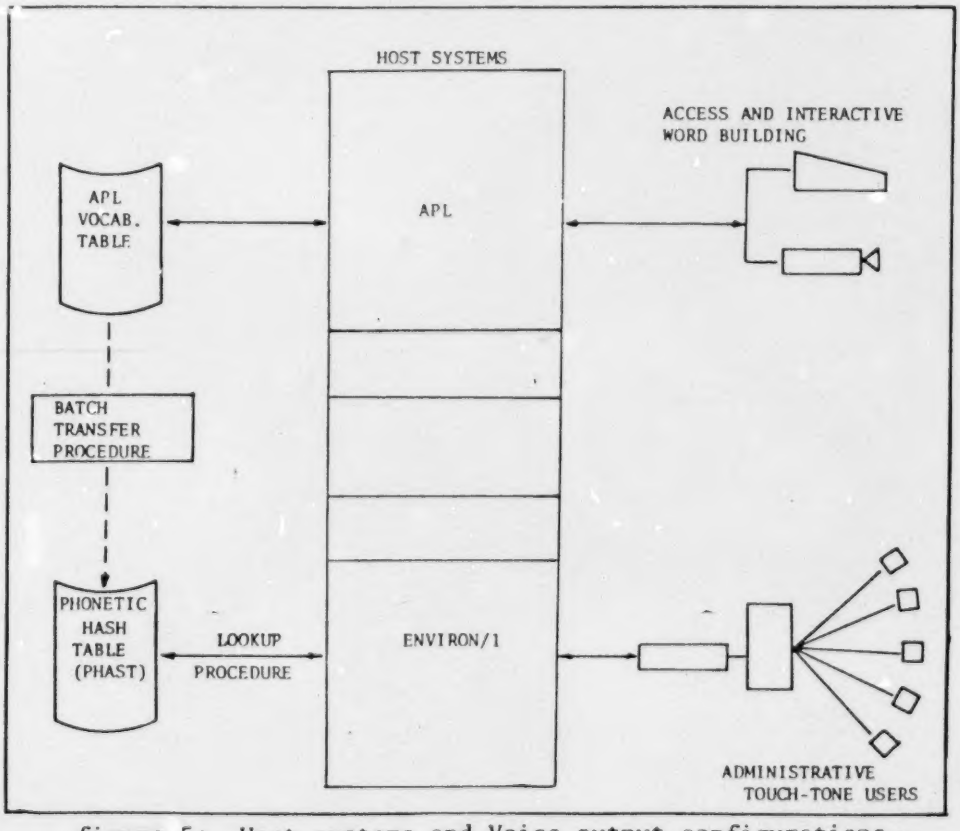


figure 5: Host systems and Voice output configurations

*APL/SV is a proprietary program product obtainable from IBM. ENVIRON/1 is a proprietary program obtainable from CINCOM Systems, Inc.

It should be noted that this scheme (figure 5) is utilized mainly for its convenience. It is not necessary to have the word-building capability singularly in APL/SV. However, the terminal-synthesizer configuration employed by APL makes the effort quite simple. You try a word, hear it, modify the phonemes and try again.

The PHAST vocabulary table is structured with a prime hash algorithm. This is an expedient organization for the random access imposed by the Lookup Procedure. The table is maintained at a load factor below 70% to enhance search performance. Our average search is accommodated by 1.8 probes. And, because of the nature of ENVIRON/1, searches for common words are often resolved in buffer store.

The rather traditional process of the Lookup Procedure is shown in the flowchart of figure 6. The application program invokes the procedure by means

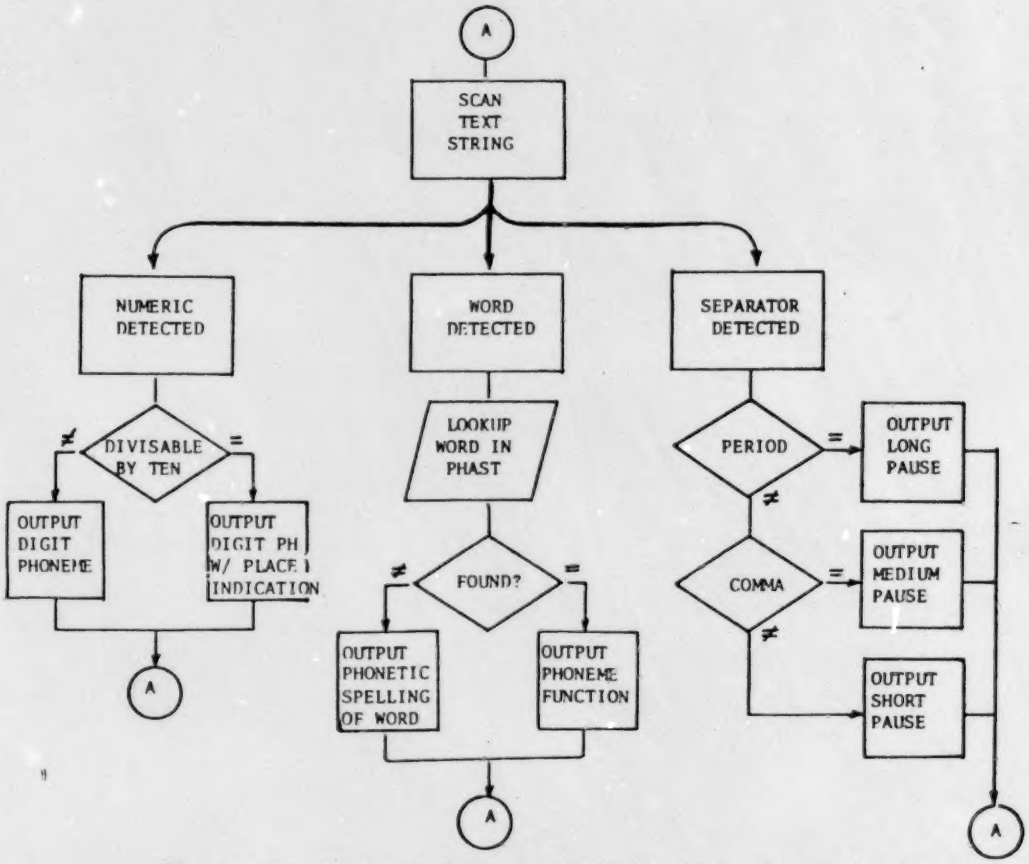


figure 6: Logical flowchart of PHAST Lookup Procedure

of a CALL and a text/sentence string parameter. Scanning determines the text elements. If a numeric is detected, it is tested to see if it is evenly divisable by 10 (modulo-10). If this is the case, the leading digit's phonemes are concatenated with the phonemes of its place indication. For instance, the numeric text 700 would be pronounced "seven" "hundred". If an application program desires place indication verbage in its output, it purposely uses this logic. For example, if the value 1,234 is sent to the Lookup Procedure as 1000 200 30 4, it would be pronounced as "one thousand" "two hundred" "thirty" and then "four". Numeric text which is not modulo-10 is pronounced digit by digit.

When a "word" or more properly a string of contiguous characters bound by separators, is detected, it is used as a search argument into the PHAST vocabulary. If the search is successful, the associated function of phonemes is output. Otherwise, the word is pronounced, or spelled, character by character.

Separators are used to issue natural pausing. As you can see (figure 6), the presence of a period effects a long pause, a comma obtains a medium pause, and all others such as hyphen, colon, blank, etc. result in a short pause.

Mention should be made at this point about an automatic pronunciation algorithm. Bell Laboritories has published accounts of a program which produces Synthetic English Speech by Rule [4]. Although the speech produced is not inflected, it is intelligible on at least 97% of running text. Implemented on a PDP11/45, it requires about 15,500 bytes It contains some 750 pronunciation rules and requires a lexicon structure for words which are determined to be exceptions to the rule. This is a promising effort, and we believe our next procedural step will follow its guidelines closely.

## 4. VOICE SYNTHESIS APPLICATIONS FOR STUDENT INFORMATION SYSTEMS

4.1 Application Design Considerations. There are several considerations which should be included in the design of any application programs which intend to use Synthesized Voice Output. An anxiety which pops into every Administrator's mind is that the "whole world" will be able to listen in to spoken output by simply lifting up their telephone. Security, therefore, should perhaps be considered before anything else. We have two levels of security checking. The first clearance is obtained after the initial connection when the User is asked "WHAT IS YOUR USER NUMBER?" He must respond with a pre-assigned number followed by a password string of digits. If his entry at this point cannot be found within our User Security Registration list, the User is asked to input the number again. If the second attempt also fails, we hang-up the connection and write a message to the computer operator indicating that an unsuccessful "sign-on" was attempted.

Assuming a good sign-on, the User is then asked to "ENTER PROGRAM NUMBER". He, of course, must know the particular number associated with the application program. After this number is entered, the User's Security Registration record is reviewed to ascertain whether he is authorized to proceed.

Two other considerations are Brevity and Clarity. The old expression "keep it short and to the point" is an ideal mandate for Voice Output messages. It will undoubtedly be the application designer's first inclination to make the voice output as "human-like" and natural as possible. But, no one wants to hear a long-winded computer. Abbreviated, precise sentences are very effective, especially when the User will hear the sentence everytime he accesses the application. Clarity, by the way, is often enhanced with good pausing between words. (When in doubt, throw in a pause.) Further clarity is also obtained by setting the Synthesizer's voice pitch to a low bass.

The application program modules should always be both interruptable and repeatable. These are User considerations. If he has heard all of a message he needs, he should have the capability of breaking the transmission. We use the '*' key on the touch-tone pad for this function The User may also wish to hear a message again in the case some interference occurred. Although we do not have a "repeat" function key, all applications are divided into brief data-text modules. Any module can be repeated simply by calling for it again.

Certain limitations should be realized if the touch-tone phone is used as a "terminal". All User input can only be numeric. This imposes quite a handicap on data-entry. There are schemes, however, where the User indicates with one series of digits that the next series of digits is supposed to be alphabetic and visa-versa, but these are cumbersome, to say the least.

Anticipating the User demand for an application is an important consideration because it will hopefully help determine the traffic increase on the equipment. Unfortunately, this evaluation will lead the designer into telephone queuing theories and their complicated analysis. But, a sure death for a new application is the User community's reaction, "yeah, sure it's great, if you can ever get past the busy signal".

4.2 Student Information Inquiry. Our first voice output application for Administrative Users was designed primarily for some 50 Counselors within the district. The initial draft of the facility was purposely kept simply as we knew the implementation effort would involve a tremendous education in the technical support area. The primary function of the application is to quickly provide the Counselor with student demographic data as well as summary information on his current activities.

The flowchart in figure 7 depicts the logical construction of the application program and its User interface. After passing a resident security procedure, control is passed to the application's main module which immediately identifies itself by outputting "This is the Student Information Program". A read is then issued to the User's terminal/telephone after the prompting word "Proceed" is spoken. At this point, the User has several options. If he wishes to terminate, he inputs (keys-in) '999' whereby control will be returned to the security procedure and subsequent opportunity to select a different application program will be offered to him. The other input options are either a student number specification, or an inquiry module code. A student number should be

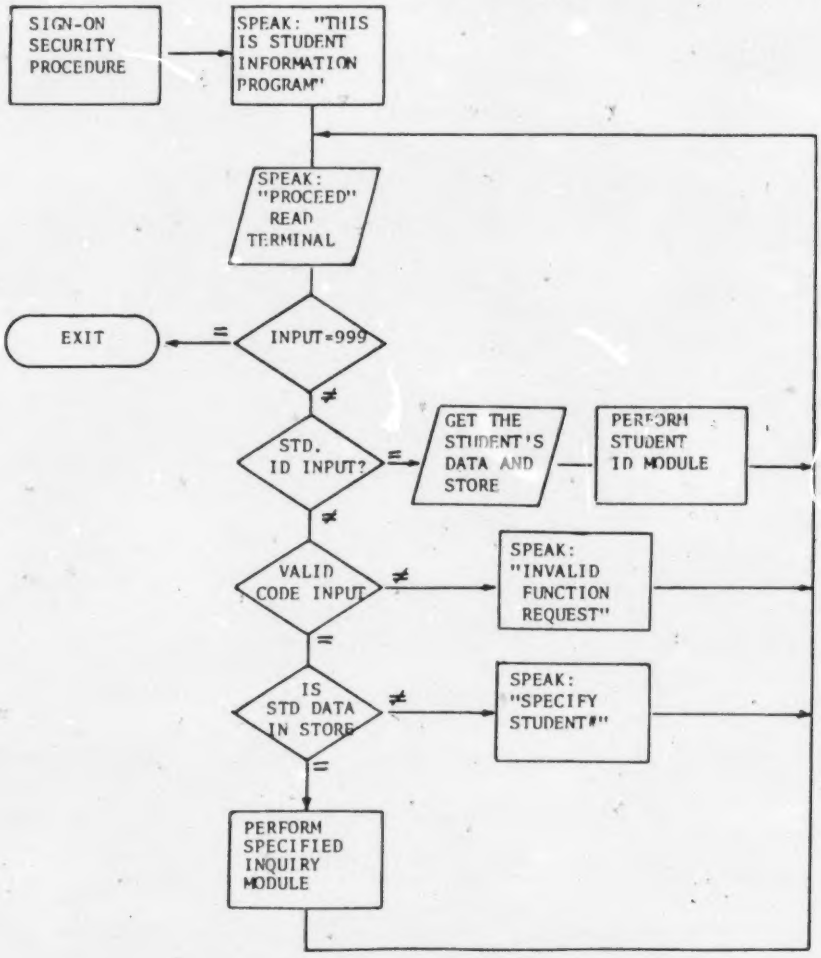


figure 7: Logical flowchart of Student Information Voice Output application program

specified first as it is the impetus for the Data Base call and the storing of the student's information for subsequent interrogation. A module which repeats the student's ID number and spells the name is performed at this point to verify that the correct student has been accessed.

Once the student's information is in storage, several brief inquiry modules may be requested, in any random order. For instance, if the User responds to "Proceed" with the digit 3 input, an associated module outputs summary information about the student's current activities/status. The text of this module is shown in figure 8. Note that variable data is easily concatenated with constant text to form one message string for the Phonetic Lookup Procedure and the Synthesizer.

INQUIRY MODULE - DIGIT 3

Spoken if Withdrawn

"STUDENT IS TOTALLY WITHDRAWN"

Spoken if Active

"STUDENT IS ACTIVELY ENROLLED IN (number) UNITS,
FOR (number) CONTACT HOURS. (HE/SHE)
IS WITHDRAWN FROM (number) UNITS.
STUDENT IS CLASSIFIED AS A (FULL,PART)
TIME (DAY, EVENING) STUDENT."

if applicable

"(HE/SHE) IS CONCURRENTLY ENROLLED IN HIGH SCHOOL."

figure 8: Output messages of inquiry module for Student's Current Enrollment Status

We currently have eight various inquiry modules which are requested by associated inputs of 1 thru 8. The effects of these modules are summarized in figure 9. The inquiry module for the digit 2 input perhaps requires clarification. Our district is composed of two colleges. Three semesters are always maintained on-line. Hence, there are six different semester data sub-sets. The User must specify (via conversational input) which college and semester he wishes to

reference. Once specified, the reference is stored with the User's security registration, and used by all subsequent Data Base calls. The User may change the reference at anytime.

INQUIRY MODULES

| TOUCH-TONE INPUT CODE | DESCRIPTION |
|---|---|
| 7 DIGIT ID | STUDENT'S RECORD IS RETRIEVED AND STORED. HIS ID NUMBER IS SPOKEN FOR VERIFICATION. HIS NAME IS SPELLED. |
| 1. | A MENU OF AVAILABLE INQUIRY CODES IS SPOKEN |
| 2. | USER IS ASKED TO INPUT COLLEGE AND YEAR SEMESTER REFERENCE. |
| 3. | STUDENT'S CURRENT ENROLLMENT STATUS IS SPOKEN. |
| 4. | STUDENT'S CUMULATIVE GPA DATA IS SPOKEN. |
| 5. | STUDENT'S MOST RECENT SEMESTER GPA DATA IS SPOKEN. |
| 6. | STUDENT'S CHARACTERISTICS (SEX, AGE, MAJOR, ETC.) IS SPOKEN. |
| 7. | STUDENT'S ADDRESS AND PHONE IS SPOKEN. |
| 8. | STUDENT'S ACADEMIC STATUS (PROBATION, DEANS LIST, ETC.) IS SPOKEN. |

figure 9: The various inquiry modules and their input codes

Upon completion or interruption of any module, control is returned to the prompting "Proceed" and a terminal/telephone read. (A module may be interrupted at anytime by depressing the * key.) This affords the User the ability to stop something he does not need to hear, or easily repeat a module which he could not hear well.

If the User forgets which input digits relates to the various modules, he may input a 1 which will result in a "Help" message that details a menu of the available modules within the application program. We plan to make this a standard for all future voice output applications. This is a more expedient reference than the User Manual ... (which for some reason can never be found!)

4.3 Other Applications.

As mentioned previously, we are able to visualize several other Voice Output Applications for our Administrative Users, as well as a number of extensions to the program just described.

One of our analysts is currently designing Voice Output interface to the on-line Budget System. The scope of this application will provide all department heads the capability to inquire about their particular budget accounts; the balances, P.O. status, encumbrances, etc. Because the district operates under a centralized accounting organization, this facility will expediate the often burdensome communication now in effect.

We are also laying out an application which will output numerous on-the-spot enrollment statistics as the On-line Registration procedure is in progress. This will include the capability to alter the OPEN or CLOSED status of a section.

Rather than supply each of our programmers with desk calculators, we are implementing a Voice Output Application for simple calculations. This will include the decimal, octal, or hexadecimal base specification.

There are many other ideas, which hopefully, the reader at this point can also visualize.

5. SUMMARY

The relative effectiveness of the touch-tone/voice output facility provided for our Administrative Users needs to be evaluated from two points of view; the User's and the Cost.

The User's appreciation for the facilities far outweighs his criticism. To begin with, he considers the "talking terminal" a dramatic advance in service provided by the Information Services department. He relates that he is able to react more immediately to voice than to any other means of communication. (Immediacy, by the way, is the real key to proper voice output information processing. [5].)

The User is also pleased that he has a sufficient number of "terminals"; effectively there is one on every desk, every office. This is to say nothing of the fact that he may use his home phone, or any other off-campus instrument. And, these "terminals" are compact and easy to operate! This is to say nothing of their dependability.

The medium is only effective, however, for applications which are limited in scope. Speech rate is considerably slower than eye-scanning. Therefore, more complex and detailed output is often better communicated in printed form.

When the User first hears his talking terminal, he will complain that it does not sound right. He is correct; it has an electro-mechanical accent which takes a few minutes to become accustomed to. It is possible to achieve extremely natural sounding messages and sentences using inflections, but it is an arduous programming task and does not lend itself to the word by word lookup-output sequence. The User will quickly adjust anyway.

The most significant amount of effectiveness is realized by the cost. If, for example, a two synthesizer configuration is procurred for $8000 and they adequately support some 80 Administrative Users, the net terminal unit cost works out to be $100 each.* And, by allowing this volume of terminals, you are really supplying an Information Service.

We fell very strongly about the possibilities of Synthesized Voice Output, and are amazed at the results we have achieved with it to date. We believe it will become a widely used medium in the near future as more Educational institutions learn of its advantages. To these installations, we of course extend our experiences and encouragement.

---

*Touch-tone pad adapters, ranging in price from $60 to $100, are available for standard telephones.

## ACKNOWLEDGEMENTS

Several individuals within the district have aided the author either in the preparation of this paper, or with the implementation of the described procedures.

-Robert Schaulis (Director, district Information Services)

-John Clark, Richard Mercer and Donald Rueter (Faculty Advisors in CAI)

-Mike Benson, Systems Programmer

-Michael O'Connor, Student Assistant

-Wiley Griener, Student Assistant

## REFERENCES

[1] J.L. Flanagan, Speech Analysis Synthesize and Perception, Springer-Verlag Berlin · Heidleberg, 1972, 204-276.

[2] D.B. Rueter, Speech Synthesize Under APL, Proceedings of the Sixth International APL Users Conference, May 1974, 585-596

[3] J.B. Eulenburg and Morteza Amir Rahimi, A Computer Terminal with Synthetic Speech Output, Behavior Research Methods & Instrumentation, 1974, Vol. 6, No. 2, 255-258.

[4] M.D. McIlroy, Synthetic English Speech by Rule, Bell Telephone Laboratories, March, 1974.

[5] All About Voice Response, Datapro 70, September, 1974.